**Twitter Analysis Project**

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Due Date (Week)** | **Notes** |
| Desk crits | 2/26 | Bring completed work to class |
| Project pinup | 3/11 | Bring printed visualizations to class |
| Mid-term presentation | 3/18 | Present in class |

**The data set**

The data set consists of five CSV files, linked primarily through the twitterUserId, which uniquely identifies every account. A complete schema diagram is provided below, and full descriptions of the fields from each file are included in the appendix.

A screenshot of a computer

Description automatically generated

**Project requirements**

The deliverables for the mid-term project will include a slide deck, visualizations, and the code to produce them. The slide deck and code will be submitted on Canvas before the start of the class on the due date, and should include the complete, working code to create your visualizations and the raw data. If you have done additional post work on the visualization in other software (e.g. Illustrator), this should be noted in the code. The in-class presentation will be 12 minutes long.

**Project Pinup**

In the project pinup, each group will bring print outs of their visualizations which they will display on the classroom walls. We will go around the room and each group will give a brief presentation of their work, including a statement of the problem you intend to address or investigate, the source of data that was used, the analyses that have been conducted, and most importantly, how the visualizations produced tell a story about the problem. After the presentation, we will have a critique session of the work, discussing elements and providing suggestions on how to improve it.

|  |  |  |
| --- | --- | --- |
| Users | | |
| Records | 127 |  |
| Field | Type | Description |
| twitterUserId | integer | Unique Twitter user ID number |
| name | varchar | Name |
| profile | varchar | Profile handle (e.g. @profile) |
| display\_name | varchar | Display name |
| location | varchar | Location |
| profile\_card | varchar | Text of user profile |
| country | varchar | Country |
| join\_date | date | Date joined |

|  |  |  |
| --- | --- | --- |
| Profile Snapshots | | |
| Records | 3999 |  |
| Field | Type | Description |
| twitterUserId | integer | Unique Twitter user ID number |
| profile | varchar | Profile handle (e.g. @profile) |
| display\_name | varchar | Display name |
| location | varchar | Location |
| profile\_card | varchar | Text of user profile |
| join\_date | date | Date joined |
| followers\_count | integer | Number of followers for the account |
| following\_count | integer | Number of accounts being followed |
| likes\_count | integer | Total number of likes received across all tweets |
| tweets\_count | integer | Total number of tweets |
| checktime | datetime | Time when check was performed |

|  |  |  |
| --- | --- | --- |
| Tweets | | |
| Records | 24216 |  |
| Field | Type | Description |
| tweetId | integer | Unique Tweet ID |
| twitterUserId | integer | Unique Twitter user ID number |
| profile | varchar | Profile handle (e.g. @profile) |
| createdEastern | datetime | Tweet creation time (Eastern) |
| createdUTC | integer | Tweet creation time (UTC timestamp) |
| favorite\_count | integer | Number of times tweet was favorited |
| favorited | bool | If tweet has been favorited |
| in\_reply\_to\_screen\_name | varchar | Profile of account being replied to |
| in\_reply\_to\_status\_id | integer | Tweet ID of tweet being replied to |
| in\_reply\_to\_user\_id | integer | Twitter user ID of account being replied to |
| is\_quote\_status | integer | Is a quote tweet |
| lang | varchar | Tweet language |
| retrieved\_on | integer | Tweet data retrieval time (UTC) |
| retweet\_count | integer | Number of times tweet was retweeted |
| retweeted | bool | If tweet has been retweeted |
| source | varchar | How the tweet was posted |
| text | varchar | Text of the tweet |
| url | varchar | URL included in tweet |

|  |  |  |
| --- | --- | --- |
| Likes | | |
| Records | 278329 |  |
| Field | Type | Description |
| tweetId | integer | Unique Tweet ID |
| twitterUserId | integer | Unique Twitter user ID number |
| liking\_twitterUserId | integer | Unique Twitter user ID number of liking account |
| liking\_name | varchar | Name of liking account |
| liking\_username | varchar | Profile handle of liking account |

|  |  |  |
| --- | --- | --- |
| Followers | | |
| Records | 127 |  |
| Field | Type | Description |
| twitterUserId | integer | Unique Twitter user ID number |
| follower\_twitterUserId | integer | Twitter user ID number of follower |
| check\_date | date | Date of following check |